TheDataWeb & DataFerrett

Bill Hazard
The DataWeb Applications Office
U.S. Census Bureau

The Census Bureau's Breakthrough Technologies

✓ **TheDataWeb** is a network of public and private databases -- a virtual data warehouse on the Internet.

✓ **DataFerrett** is a web browser that searches TheDataWeb and allows analysts to recode variables, add calculations and create custom tables.

✓ **HotReports** use DataFerrett to retrieve data from TheDataWeb to produce interactive web pages with custom tables, maps, graphs and text.

TheDataWeb Defined

An infrastructure for linking different databases into a single virtual data warehouse that supports different…

- ✓ vendors
- ✓ file structures
- ✓ data types
- ✓ file security models
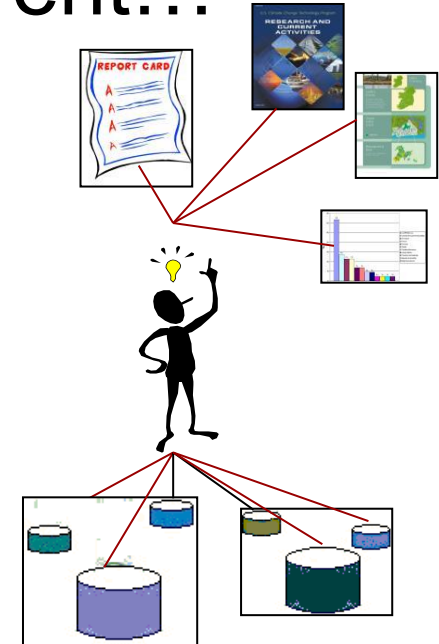
DataFerrett Defined

DataFerrett is the interface for TheDataWeb.
It allows a user to:

- ✓ Browse (search & discover) all of the datasets that are in TheDataWeb

- ✓ Select variables from multiple datasets and integrate data on-the-fly

- ✓ Produce sophisticated analyses using tables, graphs, and maps

- ✓ Create customized HotReports, or web "mash ups"

What DataFerrett Can Do

✓ Data manipulation and variable recoding

✓ Match/merge files for you (pre-merged)

✓ Frequencies, cross- and multi-dimension

   (nested) tabulations

✓ Spreadsheet formulas

✓ Mapping & graphing

✓ Publishing to HTML and PDF

Simple Cross-Tabulation

Data Visualization

Highlight spreadsheet rows or columns for:

Percent Owner Occupied Housing Units: 2000
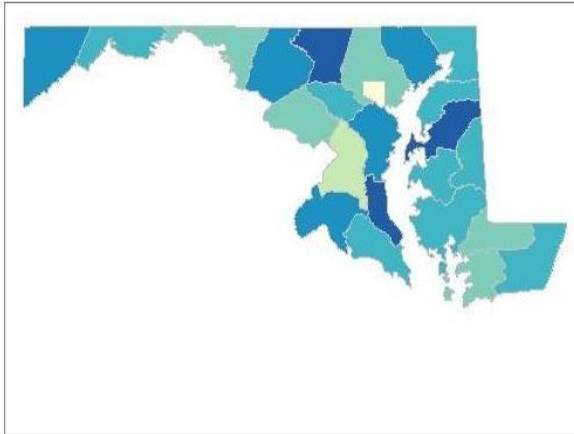
Mapping

Graphing
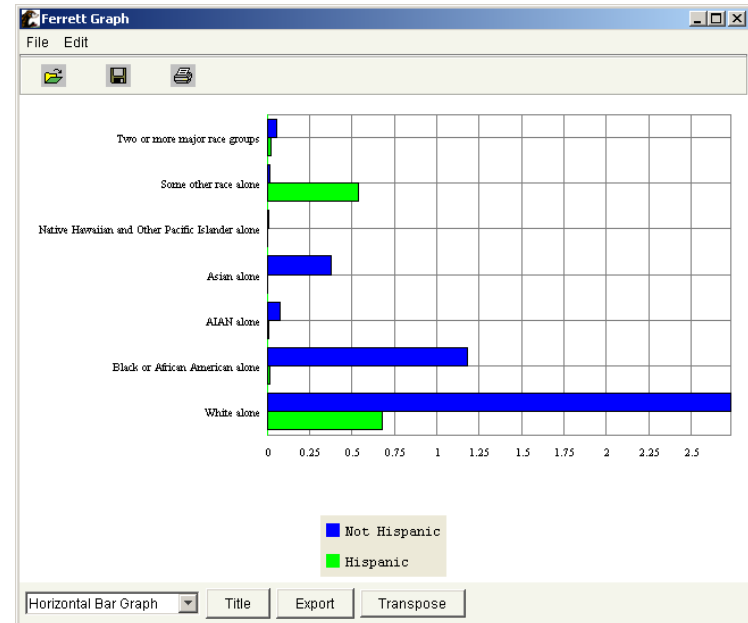
Powerful Tabulation Capabilities

Simple table layout for frequencies and trends; supporting spreadsheet math for complex analysis

- Massive Datasets
- Complex Nesting
- Formulas
- Hide columns or rows
- Use weighting variables

DataFerrett Users

Who should use DataFerrett?

✓Intended for serious data users – Power Users

  ✓Users that want to "play" with the data

  ✓Users who cannot get what they need from the pre-defined data tables

  ✓Users needing quick, yet sophisticated tabulations

United States Census Bureau

What We're Working On

- ✓ Calculating variances on-the-fly for microdata tabulations

- ✓ Advanced statistical methods – integrating the R statistical package

- ✓ Regressions

- ✓ Cluster analysis

Supports Multiple Data Types

(Needed for data integration)

Microdata: Individual transaction records, or survey response records. Data is often in multiple files (e.g. Household, Person, Diary, Geography files).

Aggregate (Macro) Data: Data that has already been tabulated. You must pick from variables or dimensions already tabulated (i.e. geography).

Timeseries: Data that can be tabulated by time periods from a cell in a table (poverty rate, or unemployment rate over time).

Longitudinal Data: follows people over time (microdata).

dataferrett.census.gov

Contact Us

**Bill Hazard –
william.g.hazard@census.gov**

**DataFerrett Help Desk –
dsd.ferrett@census.gov**

DataFerrett Features

Merging Files

- DataFerrett automatically merges files for you.

Filtering records to restrict a universe

- In SAS you use a Data Step and an "If variable X = Y;" to keep variables for analysis.
- In DataFerrett you simply click on the values of the variables that you want to include.

Creating new variables

- In SAS or SPSS, you will create a variable in a Data Step or script with a statement like "If (variable X =Y) then variable Z = variable X + 22 etc.

DataFerrett actually supports data manipulation in 3 ways.

1. Creating/saving a variable as a simple recode. This collapses the values from a single variable to more useful classes.
2. Creating a new variable from several variables using "if then else" logic.
3. Creating new aggregate variables by using the spreadsheet(spreadsheet math) to manipulate rows and columns.

Saving programs/analysis

Variables and scripted transformations are saved in two ways:

1. Saving the data basket in a DataFerrett Session File (FSF)
2. Saving the layout of a table in a DataFerrett Tabulation File (FTF) *note: this also saves the FSF associated with the table.*